UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/036,577	12/31/2001	Andrew V. Anderson	5038-175	4052

7590 09/01/2005

STEVEN P. SKABRAT
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 WILSHIRE BOULEVARD,
SEVENTH FLOOR
LOS ANGELES, CA 90025

EXAMINER	
LERNER, MARTIN	
ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 09/01/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/036,577	ANDERSON ET AL.	
	Examiner	Art Unit	
	Martin Lerner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *30 June 2005*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 to 3, 5, and 7 to 29* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 to 3, 5, and 7 to 29* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2. Claims 1 to 3, 5, 8 to 27, and 29 are rejected under 35 U.S.C. 102(b) as being anticipated by *Roberts et al.*

Regarding independent claims 1 and 11, *Roberts et al.* discloses a speech recognition system, comprising:

"at least one recognizer to produce output signals from audio input signals based at least in part on speech models and grammar files" – when step 111 detects an utterance ("audio input signals"), it causes the program to advance to step 119, which stores the token produced by step 118 in a memory buffer called TEMP_TOK; if the recognition mode has been set to TEXTMODE, step 121 causes step 123 to perform TEXTMODE recognition upon TEMP_TOK (column 8, lines 17 to 50: Figure 1: Steps 111, 121, and 123); TEXTMODE and EDITMODE use the same recognition algorithm 129, by comparing the sequence of individual frames with each of a plurality of acoustic word models 132 ("at least in part on speech models"); language model filtering may be used reflecting a probability of each word occurring in the present language context being more likely to be selected (column 8, line 51 to column 9, line 7: Figure 3); a

language model reflecting a probability of each word occurring in a current context is "a grammar file";

"a feedback module to generate feedback data" – if the recognition mode has been set to EDITMODE, step 120 causes step 122 to perform EDITMODE speech recognition on the token stored in TEMP_TOK; selection commands 125 for EDITMODE are "pick_one", "pick_two", etc., edit menu choice commands 126, such as "edit_one", "edit_two", etc., and letter commands 127, such as "starts_alpha", "starts_bravo", etc. (column 8, lines 17 to 50: Figure 1: Steps 120, 125, 126, and 127); commands for EDITMODE permit a user to provide "feedback" for correctness of speech recognition; selection commands 125, edit menu choice commands 126, and letter commands 127 are "feedback data" from a user;

"a controller adaptable to modify the speech models and the grammar files based on the feedback data to improve the performance of the at least one recognizer" – Figure 1 discloses steps of a computer program, which is "a controller"; the confirmed word is used to update the language model used by the recognition system; for each pair of words W1, W2, the probability of W2 is updated by the number of counts for how often the pair occurs as successive words in the text (column 13, lines 44 to 60: Figures 1 and 9); a language model of a probability of W2 given W1 is "a grammar file"; thus, updating a language model based upon confirming a word is equivalent to modifying "the grammar file based on the feedback data"; step 214 finds all the tokens previously stored in the tokenstore in association with the just confirmed word and builds a new acoustic model ("the speech models") for that word with those tokens; step 216 stores

this acoustic word model with the other acoustic word models (column 15, line 58 to column 16, line 6: Figure 1: Steps 214 and 216); building a new acoustic model from a confirmed word is equivalent to modifying "the speech models based on the feedback data to improved the performance of the at least one recognizer."

Regarding independent claims 16 and 25, *Roberts et al.* discloses a speech recognition method and machine-readable code, comprising:

"converting an audio input signal to an output signal" – when step 111 detects an utterance ("an audio input signal"), it causes the program to advance to step 119, which stores the token produced by step 118 in a memory buffer called TEMP_TOK; if the recognition mode has been set to TEXTMODE, step 121 causes step 123 to perform TEXTMODE recognition upon TEMP_TOK; TEXTMODE recognition is the normal recognition mode which enables the user to dictate words for inclusion in the textual output ("an output signal") of the system (column 8, lines 17 to 50: Figure 1: Steps 111, 121, and 123);

"estimating a correctness measure wherein the correctness measure expresses if the output signal is a correct representation of the audio input signal" – a score ("a correctness measure") is computed for each time aligned match between the acoustic information in each frame and the acoustic model of the node against which it is time aligned; the words with the lowest sum of distance are then selected as the best scoring words (column 8, line 58 to column 9, line 7: Figure 3: Steps 129 to 132);

"forming a feedback data element wherein the element comprises at least one of the audio input signal, the output signal, and the correctness measure" – step 174 confirms the top choice, or best scoring word, from the recognition; step 176 displays the choices from the recognition of the token just saved, with the choices displayed in order, with the top choice, or best scoring word first, and with each choice having next to it a function key number "f1" through "f9" (column 12, lines 56 to 66: Figure 1: Steps 174 and 176); confirmation of word choices by a user provides a feedback data element through selection by function keys, where feedback involves at least scoring ("the correctness measure") and confirmation of a word choice ("the output signal").

Regarding claims 2, 3, 12, 13, 15, 21, and 26, *Roberts et al.* discloses a block diagram of a computer program for coordinating output of text by speech recognition ("production of the output signals") and editing by selection commands 125, edit menu choice commands 126, and letter commands 127 ("adaptable to provide the feedback data to the recognizer") (Figure 1); the computer program is "a controller".

Regarding claims 5, 17, and 27, *Roberts et al.* discloses storing confirmed words ("the feedback data element") in SAV_TOK ("a storage"); step 214 finds all the tokens previously stored in the tokenstore in association with the just confirmed word and builds a new acoustic model ("speech models") for that word with those tokens; step 216 stores this acoustic word model with the other acoustic word models (column 15, line 58 to column 16, line 6: Figure 1: Steps 214 and 216); building a new acoustic

model from a confirmed word is equivalent to "updating speech models based on the feedback data."

Regarding claim 8, *Roberts et al.* discloses TEXTMODE recognition produces recognized text; EDITMODE recognition produces command signals (column 8, lines 17 to 50: Figure 1).

Regarding claims 9 and 22, *Roberts et al.* discloses generating feedback based upon language model filtering so that words which the language model indicates are most probable in the current context are more likely to be selected (column 9, lines 1 to 7); a language model involves "grammar files" (column 13, lines 44 to 60: Figures 1 and 9); also, each of the displayed choices are "output signals".

Regarding claim 10, *Roberts et al.* discloses generating feedback based upon user choice editing by selection commands 125, edit menu choice commands 126, and letter commands 127 (column 8, lines 37 to 50), or of function keys "f1" though "f9" (column 15, lines 27 to 40); these commands are "information received through an application programming interface".

Regarding claim 14, *Roberts et al.* discloses real time feedback as each word is recognized.

Regarding claims 18 and 29, *Roberts et al.* discloses tokens are saved only for confirmed words for adaptive speech recognition, i.e. a word that was confirmed as being correct (column 16, lines 7 to 22).

Regarding claim 19, *Roberts et al.* discloses language model filtering, where the score of a word depends upon a language model reflecting the probability of a word

occurring in the present language context ("according to a criteria") (column 9, lines 1 to 7).

Regarding claim 20, *Roberts et al.* discloses at least updating an acoustic model of a confirmed word ("updating acoustic models based on the feedback data") (column 15, line 58 to column 16, line 6).

Regarding claim 23, *Roberts et al.* discloses assigning a TEMP_TOK identifier to the token produced by an utterance for word confirmation ("as part of the feedback data element") (column 8, lines 17 to 21: Figures 1 and 2).

Regarding claim 24, *Roberts et al.* discloses confirmation of a word through language model filtering of a present language context ("identifying relevant contextual information") (column 9, lines 1 to 7).

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over *Roberts et al.* in view of *Thelen et al.*

Roberts et al. only discloses one speech recognizer, and omits multiple recognizers and a predictor to select a best performing recognizer from feedback data. However, *Thelen et al.* teaches speech recognition having parallel large vocabulary

recognition engines 331, 332, 333, where a model selector 360 is used to select at least one of the speech recognizers in dependence on a recognition context. (Column 7, Line 30 to Column 8, Line 5: Figure 3) A stated advantage is to provide a recognition system that is better capable of dealing with huge vocabularies. (Column 1, Lines 53 to 55) It would have been obvious to one having ordinary skill in the art to provide multiple speech recognizers and a selector to select a best performing recognizer based upon a recognition context as taught by *Thelen et al.* in the speech recognition system of *Roberts et al.* for the purpose of providing a recognition system that is better capable of dealing with huge vocabularies.

5. Claim 28 is rejected under 35 U.S.C. 103(a) as being unpatentable over *Roberts et al.* in view of *Ortega*.

Roberts et al. discloses a speech recognition system providing updates and adaptation of acoustic models and language models for confirmed words. Thus, *Roberts et al.* does not expressly say that audio input signals are only stored for which the correction status indicates a correction was necessary. However, *Ortega* teaches deferred correction for speech recognition systems, where a file log identifies changes to a language model and any new words added through correction. Thus, there is an advantage that a speech file can be updated on another system. (Column 1, Line 44 to Column 2, Line 6) It would have been obvious to one having ordinary skill in the art to provide a log file only for words having a correction status indicating that correction was

necessary as taught by *Ortega* in the speech recognition system of *Roberts et al.* for the purpose of permitting deferred correction on another system.

Response to Arguments

6. Applicants' arguments filed 16 February 2005 have been fully considered but they are not persuasive.

Firstly, Applicants argue that *Roberts et al.* does not disclose "a controller". Applicants state that *Roberts et al.*'s Figure 1 is a schematic block diagram of functional steps. In contrast, Applicants maintain independent claim 1 recites a controller that does not perform recognition functions because those functions are performed by the recognizer, which is another recited component. This is not persuasive.

Applicants' argument is inconsistent with the preambular statement of a speech recognition system comprising a controller of independent claim 1. Those skilled in the art would know that speech recognition is necessarily performed on a computer or processor. Thus, *Roberts et al.*'s Figure 1 is a flow chart describing steps performed on a computer or processor. *Roberts et al.* discloses speech recognition *per* se at Column 7, Lines 11 to Column 8, Line 50: Figure 2: Steps 112 to 118. Similarly, *Roberts et al.* discloses recognition *per se* at Column 8, Line 34 to Column 9, Line 7: Figure 3: Step 129. Indeed, *Roberts et al.*'s Figure 1 discloses text mode and edit mode recognition *per se* only by Steps 122 and 123. The remainder of *Roberts et al.*'s Figure 1 is directed to editing results from speech recognition and updating models. Applicants' controller is simply an element for performing a function of improving performance of a

speech recognizer by modifying speech models and grammar files. *Roberts et al.*'s Figure 1 discloses elements for performing an equivalent function of improving performance of a speech recognizer. However, *Roberts et al.*'s Figure 1, in its entirety, is concerned with a speech recognition system, just as Applicants' independent claim 1 includes a speech recognizer within a preambularly recited speech recognition system.

Secondly, Applicants argue that *Roberts et al.* does not use grammar files. Applicants contend that it is erroneous to say that a language model is equivalent to a grammar file in *Roberts et al.* Applicants state that a language model includes probabilities of a word followed by another word based upon statistics, but a grammar file is used for a command to specify what is recognizable by a recognizer. Thus, Applicants say that a grammar file is totally different from a language model. This position is traversed.

Those skilled in the art would know that the terms 'language model' and 'grammar' are frequently used as synonyms. See, for example, *Dragosh et al.*, Column 1, Lines 13 to 27, where it is stated, the term 'grammar' "may also refer generally to a statistical language model (where the model represents phrases), such as those used in language understanding systems." Similarly, *Thelen et al.*, Column 5, Line 54 to Column 6, Line 11, states, "A language model based on syntactical constraints is usually referred to as a grammar." Admittedly, there are nuances between the terms 'language model' and 'grammar', but the terms are frequently used as synonyms. As a practical matter, the terms 'language model' and 'grammar' are used interchangeably because it is significantly easier to determine what first word is adjacent to what second

word based upon statistical frequency of prior usage than it is to determine what first word is adjacent to what second word based upon grammatical rules requiring a determination of each word's part of speech (*e.g.* noun, verb, etc.).

Moreover, Applicants' Specification provides an example of a 'grammar file' that is consistent with being called a 'language model'. Page 6, Lines 3 to 20, of Applicants' Specification gives an example of "modifying a grammar file" for "call Rob" or "call Bob" by changing weightings based upon prior usage. By Applicants' own definition of a language model as being based upon a statistical probability of a word followed by another word, Applicants' grammar file for "call Rob" can be called a language model. During patent examination, the pending claims must be "given their broadest reasonable interpretation consistent with the specification." *In re Hyatt*, 211 F.3d 1367, 1372, 54 USPQ2d 1664, 1667 (Fed. Cir. 2000). Applicant always has the opportunity to amend the claims during prosecution, and broad interpretation by the examiner reduces the possibility that the claim, once issued, will be interpreted more broadly than is justified. *In re Prater*, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-51 (CCPA 1969) Claim terms are presumed to have the ordinary and customary meanings attributed to them by those of ordinary skill in the art. *Sunrace Roots Enter. Co. v. SRAM Corp.*, 336 F.3d 1298, 1302, 67 USPQ2d 1438, 1441 (Fed. Cir. 2003); *Brookhill-Wilk 1, LLC v. Intuitive Surgical, Inc.*, 334 F.3d 1294, 1298 67 USPQ2d 1132, 1136 (Fed. Cir. 2003) See MPEP 2111.01.

Thirdly, Applicants argue that *Roberts et al.* does not disclose the limitation of estimating a correctness measure. Applicants state that independent claims 16 and 25

set forth a correctness measure that is determined by a grammar file, and not by measuring an unknown input speech utterance with acoustic models, as in *Roberts et al.* Applicants say that the claimed correctness measure is not directly computed by comparing an aligned frame sequence of an unknown input utterance with acoustic word models. Applicants maintain that estimating a correctness measure is different from computing a score reflecting how close an input speech matches a word acoustic model, as taught by *Roberts et al.* This is not convincing.

Independent claims 16 and 25 do not expressly provide any limitation on how a correctness measure is estimated. Independent claims 16 and 26 only require that the correctness measure "expresses if the output signal is a correct representation of the audio input signal". Independent claims 16 and 25 do not expressly say anything about a grammar determining a correctness measure. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). Independent claims 16 and 25 only require that a feedback data element is formed from at least one of the audio input signal, the output signal, and the correctness measure. A grammar file is not recited as forming a correctness measure by either independent claim 16 or independent claim 25. *Roberts et al.* discloses a score that is equivalent to a correctness measure. Whether or not *Roberts et al.* discloses a choice list, a score is still a measure of how correctly input speech matches a word acoustic model.

Fourthly, Applicants argue, with respect to claim 7, that their claimed multiple recognizers are not specifically used to deal with a huge vocabulary, where each

recognizer is not specialized in only a portion of a vocabulary. Applicants say that each of their multiple recognizers may recognize any input utterance but with different performance. This is not persuasive.

Applicants' Specification does disclose that their recognizers 24a-24n are specialized for certain types of interactions. Page 6, Line 21 to Page 7, Line 4, of Applicants' Specification, discloses an example where recognizer 24a may be optimized for dictation and recognizer 24b may be optimized for commands. Dictation vocabularies are commonly significantly larger than command vocabularies. Similarly, *Thelen et al. ('380)* discloses a plurality of speech recognizers 331 to 333, where each recognizer is targeted to one recognition task. (Column 7, Lines 34 to 36) A model selector 360 is used to select one of the recognizers in dependence on a recognition context, or task. (Column 7, Line 63 to 66) Thus, *Thelen et al. ('380)* suggests utilizing a plurality of recognizers in a manner that is completely analogous to Applicants' claim 7.

Finally, Applicants argue that claim 28 includes storing only those audio input signals for which a correction status indicates that a correction to the output signal was necessary. By contrast, Applicants say that *Ortega* does not disclose that the log file stores audio input signals. Additionally, Applicants maintain that their audio input signals are stored for a current speech recognition system and not for another speech recognition system, as disclosed by *Ortega*.

Roberts et al. discloses saving tokens only for confirmed words. (Column 16, Lines 7 to 22) *Ortega* discloses storing a file log of corrections. (Column 1, Line 44 to

Column 2, Line 6) *Roberts et al.* may be read to store audio input signals for only those words for which correction was not necessary, as in claim 29. It is maintained that it is an obvious expedient to store audio input signals for words for which corrections were necessary once the corrections are made, given that *Roberts et al.* stores signals when correction was not necessary, and *Ortega* discloses a file log of corrections. Claim 28 is simply a complement to claim 29, and *Roberts et al.* discloses all the features of claim 29. A rejection should be evaluated for what it suggests to one having ordinary skill in the art as a whole.

Therefore, the rejections of claims 1 to 3, 5, 8 to 27, and 29 under 35 U.S.C. §102(b) as being anticipated by *Roberts et al.*, of claim 7 under 35 U.S.C. §103(a) as being unpatentable over *Roberts et al.* in view of *Thelen et al.*, and of claim 28 under 35 U.S.C. §103(a) as being unpatentable over *Roberts et al.* in view of *Ortega*, are proper.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to Applicants' disclosure.

Dragosh et al. discusses equivalence between the terms 'language model' and 'grammar'. (Column 1, Lines 13 to 27)

8. **THIS ACTION IS MADE FINAL.** Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ML
8/26/05

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER